DOCUMENT RESUME

ED 240 163 TM 840 108

AUTHOR Allen, Russell H.; Kaufman, B. Darwin
TITLE Improving Wisconsin Public School Testing Programs: With a Special Look at Pupil Competency-Based Testing.
INSTITUTION Wisconsin State Dept. of Public Instruction, Madison. Div. of Instructional Services.
REPORT NO WSDPI-Bull-3393
PUB DATE Mar 83
NOTE 33p.; For related document, see ED 209 245. Small print in tables.
PUB TYPE Reports - Descriptive (141) -- Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Accountability; Elementary Secondary Education; Minimum Competency Testing; *Program Implementation; *School Districts; Scoring; *Statewide Planning; Test Construction; *Testing Programs; Test Selection; Test Use
IDENTIFIERS *Wisconsin

ABSTRACT

This bulletin has three purposes: (1) to discuss the variety of factors which should be considered in the development of a local testing program, providing a basic understanding of what is necessary to implement and maintain a testing program; (2) to suggest several of the significant elements related to the development and maintenance of a competency-based testing program, investigating nine tasks related to establishing and maintaining such a program; and (3) to summarize some of the basic features of testing programs by presenting the results of a 1980 questionnaire designed to gather information about the nature of testing in the Wisconsin public school districts. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED240163

IMPROVING WISCONSIN PUBLIC SCHOOL TESTING PROGRAMS: WITH A SPECIAL LOOK AT PUPIL COMPETENCY-BASED TESTING

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Wisconsin Department of Public Instruction
Herbert J. Grover, State Superintendent

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. Hammatuck

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IMPROVING WISCONSIN PUBLIC SCHOOL TESTING PROGRAMS: WITH A SPECIAL LOOK AT PUPIL COMPETENCY-BASED TESTING

Russell H. Allen and B. Darwin Kaufman
Wisconsin Department of Public Instruction

Published by

Division for Instructional Services,
John Benson, Assistant Superintendent

Bureau of Assessment and Testing
Tom Stefonek, Director

March 1983

The Department of Public Instruction does not discriminate on the basis of race, sex, age, religion, handicap, or national origin.

Bulletin No. 3393

Introduction

During the past several years, professional educators have faced increased pressure to show they are meeting students' needs for quality education. The public, local boards of education, and elected officials are demanding that schools provide evidence of success. While it would be difficult to list all the factors which have contributed to this demand for accountability in education, the following reasons are frequently cited: long-term declines in student scores on some standardized tests; an increasing dropout rate; the problem of student discipline; increased costs of education; and a strongly and widely held belief that students are not learning the "basics".

The demand for accountability has prompted a few school districts to implement a testing program tied to specific instructional goals and objectives in all subjects and grades. In these districts, teaching strategies are based on these goals and objectives, and tests are administered to determine mastery. In contrast, most districts have responded with only the occasional administration of a commercial test.

This paper has three purposes: (1) to discuss the variety of factors which should be considered in the development of a local testing program, (2) to suggest several of the significant elements related to the development and maintenance of a competency-based testing program, and (3) to summarize some of the basic features of testing programs in Wisconsin public school districts.

Part I summarizes the factors associated with establishing a testing program. Although the level of detail is limited, there is sufficient information to provide the reader with a basic understanding of what is necessary to implement and maintain a testing program.

Part II looks at a specific type of testing program: competency-based testing. Nine tasks related to establishing and maintaining a CBT program are investigated.

Part III presents the results of a questionnaire sent to Wisconsin public school districts in early 1980. The questionnaire was designed to gather basic information about the nature of testing in Wisconsin public schools. Although the survey is three years old, and undoubtedly some changes have occurred, the overall pattern of testing is likely to be similar.

Appreciation is expressed to Frank Evans and Judd Harmon (Wisconsin Department of Public Instruction), Nick Dussault (Sheboygan Area School District), and members of the Bureau for Assessment and Testing for their assistance in preparing this document.

PART I

ESTABLISHING A TESTING PROGRAM

Establishing an effective testing program requires more than merely purchasing and administering a commercial test. An effective district-wide testing program should consist of at least the following seven elements:

1. Determining a purpose for testing
2. Establishing information requirements
3. Selecting or developing a test(s)
4. Administering the test(s)
5. Scoring and analyzing test results
6. Preparing and publishing a report
7. Providing for follow-up

1. Determining a Purpose for Testing

The initial step in creating a testing program is to define the purpose or purposes of testing (i.e., how the test will be utilized in making decisions). Unless there is a clear statement of purpose, the existing program will not be focused and may be of little value. Listed below are nine purposes for testing students:

1. identification of curricular strengths and weaknesses
2. promotion or graduation
3. placement of pupils into groups or programs
4. diagnosis of individual student strengths and weaknesses
5. research and development
6. guidance and counseling
7. evaluation of teachers
8. comparison of students to other students
9. student assessment

It is important to emphasize that each of the eight purposes is associated with the need to make decisions. For example, the purpose of identifying curricular strengths or weaknesses might be to decide if changes in the curriculum are necessary, and, if so, where they might best be made. Likewise, the purpose of assessing student strengths and weaknesses may be to select students who need remedial help.

Shown on the next page is a summary table developed by Anderson, Stiggins, and Hiscox, who suggest there are three basic purposes for testing students:

1. to make instructional management decisions,
2. to make decisions about screening students,
3. to make program decisions.[1]

---

[1]Beverly L. Anderson, Richard J. Stiggins, and Suzanne B. Hiscox, *Guidelines for Selecting Basic Skills and Life Skills Tests*, Clearinghouse for Applied Performance Testing (Portland, Oregon: 1980), p. 4

The Anderson, Stiggins, and Hiscox categorization is of particular value because it describes the relationships among general purposes (assessment context), related decisions, necessary data, and appropriate decision-makers. As a case in point, diagnosis of student strengths and weaknesses is one of the decision areas listed under Instructional Management. Because the purpose of testing in this case is to make decisions about individual students, any test which is adopted must provide information about specific skills and/or subject matter understanding.

Summary of Various Purposes for Testing

Assessment Context	Type of Decisions	Type of Data Needed	DECISION MAKERS					
			Students	Parents	Teachers	Admin.	Counselors	Public
Instructional Management Decisions								
Diagnosis	Decide students' strengths and weaknesses on specific skills	Individual student data on level of development of specific skills	✓		✓			
Placement	Place student into next most appropriate level of instruction	Scores that place students on relevant knowledge or skill continuum			✓	✓	✓	
Guidance & Counseling	Decide probability of success and satisfaction in given program of educational or vocational development	Data reflecting level of educational development of individual student relative to other students	✓	✓			✓	
Student Screening Decisions								
Selection	Decide which students to be selected into or out of a program	Data that rank order individual students on relevant knowledge or skill scale			✓	✓	✓	
Certification	Determine mastery or non-mastery of specified body of knowledge or set of skills	Data reflecting individual student mastery of specified body of knowledge or set of skills			✓	✓		✓
Programmatic Decisions								
Survey Assessment	Make educational policy decisions; determine educational development of student group	Group achievement data gathered cyclically to show trends				✓		✓
Formative Evaluation	Decide program components in need of modifation	Interim and final program outcomes attained and not attained by participating students considered as a group			✓	✓		
Summative Evaluation	Determine if program is to be adopted, expanded, or discontinued	Program outcomes attained and not attained by participating students considered as a group			✓	✓		✓

In the important initial step of defining purpose, there are two suggestions. First, there should be a cooperative effort among those who will be required to make decisions based on the test results. This will probably include teachers, administrators, curriculum specialists, counselors, psychologists, and evaluation specialists. Second, it should be recognized that any testing program must be limited in scope. It is unrealistic to propose a testing program which will answer all possible questions. Thus, there is the need to be selective throughout the process of establishing a testing program.

The discussion of purpose offered in this paper is not meant to be exhaustive. Rather, it is intended to alert the reader to the variety of reasons for testing and to the critical need to define the intended purpose or purposes for testing.

2. Establishing Information Requirements

After determining the purpose(s) for testing, the next step is to identify the type of information needed to make the desired decisions. The information provided by any testing program can be classified into one or more of three categories or domains:

1. The affective domain, which refers to attitudes, feelings, interests and values.[2]
2. The psychomotor domain, which refers to those skills involving neuromuscular coordination. Handwriting skills and athletic skills represent two types of psychomotor skills.
3. Cognitive skills, which are identified in a variety of ways. They are usually classified along a continuum. Knowledge (that is, knowledge of facts, rules, and sequences) is viewed as a "lower" order cognitive skill, while "higher" order skills include the ability to classify, to recognize relationships, to analyze, to synthesize, to evaluate, etc.

If, for example, the purpose for testing is to make a decision about the placement of students into academic classes of differing ability levels, it will be necessary to collect data which yield information about how students perform on specific cognitive objectives or skills. On the other hand, if a program has explicit affective goals, and it is decided to determine how well those goals have been reached, it will be necessary to develop or purchase valid and reliable measures of student attitudes, feelings, values, or interests.

---

[2]For a discussion of attitudes, see the following: Harry C. Triandis, ***Attitude and Attitude Change***, John Wiley and Sons (New York: 1971); ***Attitudes Toward School, K-12***, Instructional Objectives Exchange, P. O. Box 24095, Los Angeles; ***Attitudes Related to Tolerance***, Instructional Objectives Exchange, P. O. Box 24095, Los Angeles; and ***How to Measure Attitudes*** by Marlene E. Henerson, Lynn Lyons Morris, and Carol Taylor Fitz-Gibbon, Sage Publications (Beverly Hills: 1978).

As an example, suppose that one of the purposes for testing is to determine how well students have mastered the cognitive objectives of a given course. If teachers have not previously identified objectives, or if the existing objectives are judged inadequate for some reason (perhaps they are stated in extremely general terms), it will be necessary to develop a list of instructional objectives.

In carrying out this task, two steps are involved:

First, it will be necessary to state all objectives in specific terms. One way to determine whether or not an objective is stated clearly and explicit is to ask several subject matter experts to describe the types of test questions which measure the objectives. If there is a reasonable level of agreement among these individuals, then it is probably safe to assume the objective is well written.

Second, it may be necessary to limit the number of objectives to be assessed, simply because it may be unrealistic to test all possible learner outcomes. One approach for limiting the number is to have teachers rank them in importance. This way they will test only the objectives judged to be essential.

Another way to choose (limit) objectives for purposes of testing is on the basis of their position in a curricular hierarchy. For example, multiplication of whole numbers requires mastery of a number of prerequisite skills and is itself prerequisite to other higher level skills. For these two reasons, it would be a reasonable objective to test.

These two steps usually are difficult and time consuming because it is often difficult to get teachers to agree about the instructional objectives to test, the exact wording of the objectives, the level of cognitive understanding, and what they consider to be satisfactory and/or exemplary performance.

3. Selecting or Developing a Test

After identifying both the purposes for testing and the information requirements of the testing program, selection or construction of appropriate test instruments is the next step. There are two types of tests from which to choose:

a. Norm-Referenced
b. Objective-Referenced

Within the category of objective-referenced tests, there are four options:

1. Existing objective-referenced tests
2. Customized objective-referenced tests

3. Item bank objective-referenced tests
4. District developed objective-referenced tests

A. Norm-Referenced Tests

If the chosen instrument is norm-referenced, it is essential that the basic characteristics of these tests be understood. A norm-referenced test is designed to sample a broad range of a subject matter area, but such tests do not typically present information about performance on specific instructional objectives. Norm-referenced tests do "sort" or spread students along a continuum or scale. The score a student receives on a norm-referenced test has meaning as it relates to the performances of other students. Thus, a percentile score of 65 means that a student has scored the same or higher than 65 percent of the students who took the test when it was originally normed. It also should be recognized that students who attain the same percentile scores (e.g., both at the 65th percentile) need not have the same strengths and weaknesses. Each student's scores may represent a different constellation of skills and knowledge.

Typically, norm groups are state or national samples. Information about how a student (or group of students) compares to a norm group can provide a general idea of performance. However, such tests are not suitable for identifying specific strengths or weaknesses of students or for making judgments about the strengths or weaknesses of a program.

Popham makes the following observation about two basic weaknesses of a norm-referenced test.[3]

> The first prominent defect of norm-referenced tests stems from their primary function, that is, to permit comparisons of an individual's performance with that of a norm group . . . . In order to permit an individual's performance to be referenced to the performance of a norm group, there must be a reasonable degree of response variance in scores on the test. In other words, to make fine-grained comparisons (e.g., Mary is at the 76th percentile; John is at the 74th), the test must produce a reasonable spread of scores. Test items which do not contribute substantially to the production of score variance are either modified or eliminated. . . . An ideal test item for maximizing response variance is one which only 50% of the test takers answer correctly. An item answered correctly by 80% or more of the students is typically jettisoned, since it doesn't do its share in spreading learners out. This means that in many cases the very items which deal with the concepts or skills considered most important to a field

[3] W. James Popham, "An Approaching Peril: Cloud-Referenced Tests," Phi Delta Kappan, May 1974, Vol. 55, No. 9, pp. 614-615.

will, over time, be systematically eliminated from a norm-referenced achievement test. After several revisions, in fact, a standardized achievement test looks and functions very much like a standardized intelligence test. Such tests are relatively insensitive to instruction.

A second basic problem with norm-referenced tests is that they are developed by test production firms which must service an entire nation, a nation with very diverse curricular preferences. If the test agencies spelled out, in unambiguous detail, exactly what it is their tests measure, many educators would find that the test's emphases did not match the local curriculum. They wouldn't buy the test. But test publishers subsist on the sales of their tests. They get out of this bind by describing their tests in very general terms, thereby picking up the classic Rorschach dividend of allowing people to see what they want to in an ink blot. This practice results in many educators using tests which are, in reality, not consonant with local curricular emphases. Mismatched measurement and instruction yields misleading data and spurious conclusions.

In summary, norm-referenced tests are useful to screen students for purposes of placement or guidance. They also allow one to make comparisons with a norm group. Furthermore, they can provide an indication of how well students are able to answer questions representative of the broader range of topics possible in the curricular area. However, norm-referenced tests generally are inappropriate for purposes of diagnosing student strengths or weaknesses or for judging whether or not a program effectively is achieving its instructional objectives. In order to use a norm-referenced test for these purposes, it would be necessary to study in detail all test objectives and related items, and then to conclude that the test is, in fact, an appropriate instrument for a given set of objectives.[4]

B. Objective-Referenced Tests

An objective-referenced test is designed to gather information about student performance as it relates to specific instructional objectives.[5] This means that the test items are based on clearly

---

[4]Typically, norm-referenced tests will provide a list of objectives and questions keyed to those objectives. However, keep in mind that the primary intent of a norm-referenced test is to "sort" students, not to measure student performances on specific skills or objectives. Such global tests usually do not contain a sufficient number of items per objective to enable reliable decisions to be made abouteducational achievement.

[5]There is often confusion regarding the difference between an objective-referenced and criterion-referenced test. The basic difference is that the criterion-referenced test, as well as being objective-referenced, also has pre-established levels of "acceptable" or "mastery" learning. An objective-referenced test does not have performance standards.

defined objectives, and that test results are interpretable in terms of attainment or nonattainment of the objectives. Thus, in addition to providing an overall test score, an objective-referenced test provides performance data on each test objective. If it is decided to use an objective-referenced test, there are four options from which to choose. Each is discussed briefly.

1. Existing Objective-Referenced Tests

One approach is to use an existing objective-referenced test. An important factor in choosing this type of test is the degree to which the test objectives match your course objectives. In the absence of a perfect match, the test may still be of use if only scores on selected objectives are considered. For example, if two of the test's ten objectives are not stressed in the curriculum, the student scores on these objectives can be disregarded in analysis and reporting.

The test publisher should be able to demonstrate that a thorough process of defining objectives and developing items was used. It is particularly important that the publisher be able to demonstrate that the items are valid measures of the stated objectives.

An objective-referenced test may also provide norms (for example, state averages). However, this feature is secondary to its primary purpose: to provide information about student performance on a battery of objectives.

2. Customized Objective-Referenced Tests

If district needs cannot be met by published tests and it is decided to develop tests, one option is to contract for customized objective-referenced tests. The usual procedure is to develop a list of objectives and present them to a testing company. In turn, the test publisher will provide appropriate test questions for each objective. If it is decided to use this approach, consult the contractor in advance because specific procedures differ from company to company.

There are at least two disadvantages associated with the use of customized objective-referenced tests. First, this type of service can be quite expensive. Second, because the test is designed to meet specific needs, it will not be possible to compare student scores with another group. If comparisons are an important component of a testing program, it will be necessary to use a test which has been widely administered and has norms associated with it.

3. Item Bank Tests

A second option for developing tests is to use an item bank. Ideally, an item bank refers to a computerized collection of items which are readily accessible to potential users. Such a system allows a user to select items from a variety of choices keyed to established

objectives. Because the process of test construction is computerized, a functioning item bank allows teachers to develop tests with relative ease.

In addition to completely developed item banks, there are numerous collections of items which may be of use for test development. To use such collections, users must first match their objectives to those of the particular collection and then access the items keyed to the objectives of interest. The process in this instance is not computerized. As a result, developing a test will probably be more tedious and time consuming.

4. District-Developed Tests

A third option is to have district personnel develop the tests. If this approach is selected, it is recommended that a basic textbook in tests and measurement be consulted.[6] Usually there is a section which deals with the principles of test construction. It is not practical to discuss the details of developing a test in a paper of this length.

Locally developed tests have great appeal because they can be designed to measure a district's specific instructional objectives. However, a major consideration is the significant amount of time and effort required to design a valid and reliable test instrument. For example, each test produced by the Wisconsin Pupil Assessment Program requires approximately two years to develop and involves between 50 and 100 persons. Time and effort are needed to develop objectives, write and revise test items, field test items, and pilot test items on a statewide sample.

Regardless of which test option is chosen, there are two related questions which require answers:

1. Who is to make the choice of a test or test items?
2. What criteria should be considered in making the choice?

As mentioned previously, decision makers should be involved in choosing a test or test items. Without question, the judgment of teachers is needed because they have the subject-matter expertise which is crucial in reaching a final decision. However, there is another equally important reason for their involvement. If the testing program is to be effective, it must have the cooperation of teachers whose students are to be tested.

---

[6]For example, see Robert Ebel, *Essentials of Educational Measurement*, Prentice-Hall (Englewood Cliffs, New Jersey: 1979); and Kenneth D. Hopkins and Julian C. Stanley, *Educational and Psychological Measurement and Evaluation*, Prentice Hall (Englewood Cliffs, New Jersey: 1981)

The answer to the second question, "What criteria should be considered in making the choice?," is more complex. The most important criterion must be whether or not the test will produce the needed information.

There are several additional criteria which might be used to judge potential tests, whether purchased or developed. These criteria are particularly applicable if it is decided to purchase a test from an outside source.

The following are among the other most important criteria:

1. *Validity:* This refers to how well the test measures the established objectives. For an externally developed test, note carefully how the test publisher addresses the question of validity.

2. *Reliability:* Scores on a reliable test are accurate and consistent. This means that students with similar skill levels would obtain similar scores. It also means that if it were possible to retest the same students, their second scores would be very close to their first scores if they had not received training between testing sessions.

   Almost all test publishers provide information about the reliability of their tests. Also, most scoring services provide statistics about the reliability of the set of test results analyzed. Typical statistics are the standard error of measurement and a measure of internal consistency, such as the Kuder-Richardson formula 21 coefficient or the Cronbach Alpha coefficient.

3. *Ease of Administration:* Are the directions provided for both the students and those who administer the test clear? If not, the results may be invalid and unreliable.

4. *Administration Time:* Is the time required to test students reasonable, or is an inordinately large or small amount of time required?

5. *Scoring Service:* What types of scoring services are available? What types of reports are available? What is the turn-around time? Does the scoring service provide the kinds of item data and test statistics needed?

6. *Costs:* What are the costs of test booklets, technical manuals, answer sheets, as well as the basic and optional scoring services?

7. *Special Features:* Does the test have any features which are particularly attractive? What are those features? How might they be used? Do they coincide with the original purpose for testing? What do they cost?

4. Administering the Test

If a test is purchased, carefully read the administration instructions provided by the test publisher. In addition to these instructions, there are some general rules which should be followed in any testing situation. Among these are the following:

1. Be certain that all students understand the purpose for testing as well as the mechanics of the specific testing situation.
2. Organize all materials prior to the testing period in order to use time effectively.
3. Provide a proper testing environment: adequate lighting and ventilation, minimal distractions, etc.
4. Conduct multi-part tests or lengthy tests in separate testing sessions.
5. Avoid test administration on days before vacation or at times prior to special school events (e.g., the hour before an all-school assembly would probably be a poor time to test).
6. Make some provision for students who are absent to take the test.

5. Scoring and Analyzing Test Results

It is not possible to adequately discuss scoring or analysis of test results in a brief paper. Scoring and analysis, however, represent essential components of a testing program. These aspects of a testing program involve more than making arrangements for scoring and then tabulating the test scores into summary form. The data need to be organized and studied in a systematic manner. It is important that a person with training and experience in the analysis of test results be involved in this phase of a testing program.

It is also advisable that additional kinds of information be used in the analysis of test results. For example, student scores on other tests or consultations with teachers, as well as with the person who administered the test, may provide information which is of value in understanding the results of a specific test. Above all, it should be recognized that no single test measures all that students learn or experience. In reaching any conclusions or making any recommendations, this point must be kept in mind.

6. Preparing and Publishing a Report

A report should address five essential topics[7]

[7] An excellent summary of reporting test results can be found in the *Program Evaluation Kit* by Lynn Lyons Morris and Carol Taylor Fitz-Gibbon, op. cit. See the publication entitled *How to Present an Evaluation Report*.

1. What are the purposes for testing?
2. Who was tested, what areas were tested, and when did testing occur?
3. Why was a particular published test chosen, or, if tests were locally developed, what was the process of developing them?
4. What do the results show?
5. What are the conclusions and recommendations?

While all reports should deal with each of the above, the specific emphasis and amount of detail will vary, depending on the target audience. Most likely, it will be necessary to prepare different kinds of reports, simply because the information desired by one group is unlikely to be the same as that desired by another. (As an example, think of how a report to the school board or teachers would differ from that developed for a news release).

In preparing an oral or written report, the following are recommended:

1. Keep your audience in mind at all times. Limit the presentation by focusing it towards the kind of information they can use. Emphasize the important points. Don't present too much detail.
2. Think of ways to make your report interesting.
3. If the report is oral, allow time for questions. If the report is written, encourage feedback.
4. Use graphs, charts, tables, and other visual aids.
5. Exercise caution in making comparisons among schools or classes on test scores. If comparisons are to be included in the report, recognize that there are considerable risks involved.
6. Emphasize the fact that a test measures only part of what takes place on a daily basis. Under no conditions should the scores from a single test (or even scores from a series of tests) be used to judge the total educational program.

Analysis of test results should provide some ideas for improvement; these ideas may be included in the report. Some of the ideas will have been considered during planning, when decisions to be based on findings were discussed. Others may occur to the report writer after examination of the data. Recommendations can be included within the final report, or if preferred, they can be presented separately. Recommendations might include initiating particular teaching strategies, holding inservice programs, acquiring new materials and textbooks, giving additional emphasis to selected topics, modifying curriculum and objectives, and providing additional help to certain students. Making decisions of this type

usually requires a background in such areas as learning theory, curriculum, and testing. Persons with expertise in any or all of these areas represent ideal resources for this stage of a testing program.

7. Providing for Follow-up

The basic reason for initiating a testing program of any type is to gather information to make educational decisions. If the test data provide the necessary information, it should be possible to make decisions which are effective. What is needed as a final step in a testing program, then, is follow-up, in which it is determined whether or not the decisions or recommendations have had the desired effect. Additional testing and discussions with teachers and students represent two ways to determine whether or not the changes have been of value. In effect, the follow-up is the first step in planning for the next cycle of testing.

SOME RECOMMENDATIONS FOR TESTING PROGRAMS

There is no single model or approach to establishing a strong testing program. However, listed below are six recommendations which should be helpful in establishing one which is both effective and received positively.

1. Testing should be conducted in a positive manner. A testing program should be intended to improve the quality of both educational programs and instruction. If weaknesses are indicated, emphasis should be on correction, not censure.

2. All testing efforts should be realistic in scope. It is not possible to measure all that a student experiences in school. What this means is that there is a need to be selective throughout the entire process of testing.

3. It should be the goal of each subject area to build a testing component into its instructional program. For example, a final examination in a course can be designed not only to provide a basis for student grades, but also to provide teachers with helpful information to improve instruction and to give district-wide programmatic information.

4. It is important to have consensus in all testing efforts. Teachers, administrators, board members, and parents should all participate in identifying course objectives, developing or selecting tests, analyzing data, and recommending changes.

5. If existing tests are used, the particular features of those tests should be well understood. Above all, the tests selected should reflect the district's instructional objectives. If they do not, the test scores will certainly be invalid and may be misleading.

6. Test information collected regularly over a period of time is desirable. For this reason, it is recommended that a testing program be carried out over a period of years. A one-time testing effort is of relatively little value.

7. If existing tests which provide norms are used to obtain comparative data (e.g., state or national comparisons), it is advisable to administer the tests at the same time during the school year in which the tests were normed. To do otherwise may produce invalid results.

PART II

ESTABLISHING AND MAINTAINING A COMPETENCY-BASED TESTING PROGRAM

This section of the paper offers a general look at several of the most important tasks related to establishing and maintaining a specific type of testing program: competency-based testing (CBT). It is assumed the reader is familiar with part I of this pamphlet, in which there is a discussion of the basic features of an effective testing program. This earlier material is relevant, and reference is made to it at several points in the ensuing discussion.

At the very least, it would appear that an effective competency-based testing program must address nine topics or areas of concern. Failure to deal with any one of these topics may jeopardize the overall effectiveness of a CBT program. This is especially true if the competency program has promotion or graduation requirements. In this case, it is essential that students be notified of all competency requirements, that the competencies be addressed effectively in the district's curriculum and instruction, that students be given adequate notice that graduation is tied to mastery of certain competencies (five years is probably adequate), and that students be given remedial help should they fail to master the competency test(s).

The nine topics are as follows:

1. Planning and Organizational Activities
2. Identification of Competencies
3. Test Development/Adoption
4. Standard Setting
5. Test Administration and Scoring
6. Record Keeping
7. Reporting Results
8. Remediation
9. Evaluation

Each of the topics will be discussed; alternatives will be noted; and when appropriate, recommendations will be made.

1. Planning and Organizational Activities

The experiences of school districts and state education agencies which have established competency-based testing programs remind us that there are numerous activities which must be addressed prior to actual program implementation. The following represent several of the more salient ones:

1. An overall plan for program development, with timelines, should be established at a very early stage. That plan should address each of the nine points listed above.

2. Individuals who are to assume important roles in various stages of program development should be identified. It is recommended that one person be assigned primary responsibility.

3. Committees will likely assume important roles in the competency program. It is recommended that a CBT task force committee, assigned overall program responsibility, be created and meet on a regular basis. The committee should include persons from the district whose representation is crucial to success: district superintendent, building principals, curriculum director, subject matter specialists, teachers, district testing coodinator, etc. Representation of members of the general public also is recommended. Other committees (and potential committee members) should be identified as well. These committees will likely be created on an ad hoc basis and have a limited focus.

4. A plan for publicizing the program (both within and outside the school system) should be developed. The public and school staff should be informed about the goals and major activities of the program.

5. Numerous policy decisions will have to be made. Representative examples follow:

   -- at which grade levels and in which subjects are competencies to be developed?

   -- what provision(s) is to be made for special students? for transfer students? for limited English proficiency students?

   -- what is to be the level of public involvement in the program?

   -- what kinds of written materials are to be developed? Who is to have primary responsibility for developing these materials?

   -- what type of technical assistance will be needed? What is to be the source of that assistance?

- -- is demonstration of competency to be a requirement for promotion or graduation? Is it to be one of several requirements? If so, what are the other requirements?
- -- how is student competency to be judged? Are paper and pencil tests to be used exclusively, or are other measures of performance to be used?
- -- what are the estimated program costs? Are there sufficient funds budgeted to competency-based testing?
- -- is there adequate space (which is secure) for storing, packaging, and distributing CBT materials ?

These points represent only a few of the many issues which should be considered. Others will become apparent as the plan for the program is developed and implemented.

2. Identification Of Competencies

Identification of competencies which students are expected to master is crucial to overall program success. In the process of defining competencies, it must be demonstrated that there exists a linkage among curriculum, instruction, and competency requirements. In other words, it must be shown that the competencies for which students are held accountable are addressed adequately in the district's curriculum and instruction.

Prior to beginning the task of defining competencies, it is recommended that several rules or guidelines for competency development be established. If this is not done, it is possible that the end product (in this instance, a list of competency requirements) will consist of a mixture of unrelated and ill-defined learner expectations.

Listed below are five points which should be helpful in identifying competencies.[8]

1. A competency represents a statement of intended (expected) student outcome.
2. The outcome is stated in clear and concise language.
3. The outcome is measurable, using plausible and practical techniques.
4. The outcome refers to a relatively general domain of behavior. (This rule seems reasonable. However, it would appear that competencies for the elementary grades would be more limited and specific.)

[8]These points are based on ones presented by E. Allen Schenck, *A Guide to Identifying High School Graduation Competencies*, Northwest Regional Educational Laboratory, (Portland, Oregon: 1978), pp. 5-6.

5. The outcome describes knowledges, skills, or attitudes necessary to function in adult life roles, or the outcome describes knowledges, skills, or attitudes fundamental to future learning.
6. The outcomes are stated in a uniform format: "The student can . . . (transitive verb) . . . predicate."

In summary, a competency can be defined as a clearly written statement of expected student outcome. The outcome statement may describe a skill, knowledge, or attitude necessary to function as an adult, or it may describe a skill, knowledge, or attitude fundamental to future learning. The competency should refer to a relatively general domain of student behaviors and should be measurable. The competency may or may not be linked to promotion or graduation requirements.

Several specific questions related to the process of defining competencies must be considered:

a. How many competencies will be required?
b. What specific process will be used to develop competency requirements?
d. How many individuals will participate in the process of competency identification? Which persons should be involved?
e. What level of public involvement is anticipated? (Public involvement is recommended, for it enhances community understanding and support).

3. Test Adoption/Development

Student attainment of competencies need not be judged exclusively on the basis of test results. Examples of other types of evidence may include completion of a product, such as a poem, artwork, or term paper, or performance of a speech or athletic routine.

However, for most districts, tests will be the primary tool for determining competence. Pages 5-10 of this pamphlet are a useful resource regarding the various types of tests which are available and the process of test selection or development.

4. Standard Setting

Standard setting represents an essential component of a competency-based testing program. There are at least two reasons for setting performance standards. The first is rather obvious: there is a need to distinguish between students who demonstrate competency and students who fail to show competency. There is a second important reason. Performance standards can serve to make test scores more understandable to the general public.

It needs to be emphasized that although there are several methods which can be used to establish standards, each requires human judgment. There are no statistical procedures which allow one to avoid subjectivity in the standard setting process.[9]

At the time a specific method of standard-setting is selected, it will be necessary to decide whether there is to be an overall test standard (e.g., the student must answer correctly 85% of the items) or separate standards for each competency. The latter is recommended, for it allows evaluation of specific strengths and weaknesses of students. This information is essential for planning remediation procedures. (This means that each competency must be measured by an adequate number of items).

Still, even though mastery for each competency is recommended, it may be that an overall test mastery score is desired. These two types of analysis can be coupled effectively. For example, competency of mathematics could be demonstrated by passing 80 percent of the mathematics items and also by mastering 80% of the mathematics competencies.

Practitioners who have had experience in setting performance standards suggest several guidelines:

1. All procedures to set standards require human judgment. From a practical standpoint, the procedure used must be thorough and systematic. Because test items vary in difficulty (even for those measuring the same competency), it is advised that each test item be looked at carefully.

2. Standard setting should involve the efforts of several individuals, who are representative of significant interest groups. It is not recommended that only educators determine performance standards.

3. The standard setting process should be documented and made available for public scrutiny.

4. Periodically, standards should be reviewed, and if necessary, revised. In short, this means that standard setting never can be a one-time exercise. Hambleton suggests that if a testing program is in its early stages, it may be advisable to err on the side of setting lower standards, since it will be easier to raise standards than to lower them.[10]

---

[9] A useful publication has been developed by the Massachusetts Department of Education: Standards Setting Manual, Massachusetts Department of Education: Boston, Massachusetts, 1981). Another useful publication has been developed by Hambleton and Powell: Ronald K. Hambleton and Sally Powell, ***Suggestions for Improving the Standard--Setting Process***, Laboratory of Psychometric and Evaluative Research, Report No. 118, (Amherst, Massachusetts: School of Education, University of Massachusetts.)

[10] Ibid., page 17

5. When it is available, test data should be used in setting standards. Results from earlier testing can provide valuable information regarding the effects of setting standards at specific levels.

6. One must always be on guard against allowing standards, which may represent minimal expectations, from being perceived as learner maximums.

7. It needs to be recognized that when standards are developed there will be students who will fall on the borderline, being neither a clear pass nor clear fail. Keep in mind that any performance standard requires human judgment and that all testing involves errors of measurement.

8. Those who contemplate using competency test results as the basis for graduation decisions should recognize that standard setting procedures likely will be a focus of any legal challenge.

5. Test Administration And Scoring

Pages 10-11 of this booklet deal with selected topics related to administering and scoring tests. The ideas presented have application to competency-based testing.

Test security is a primary concern in all testing situations. However, it may be even more important in competency-based testing because of the potential impact (promotion/graduation, remediation) of student performance on the tests.

A secure storage area is a must. But equally important is the need for rigorous accounting procedures. Hanna suggests that a structured method of distributing and collecting materials is imperative to maintaining test security.[11] He suggests the following:

> Another of the less challenging aspects of supervising a test administration is counting testing materials repeatedly. When materials arrive from the publisher, they should be counted accurately enough to ensure a sufficient supply.
>
> When these are packaged for local schools they should be counted and recounted, preferably by a different person. When they arrive at the school, test booklets and answer-sheets should be counted by the local coordinator before storing. When they are distributed by the school coordinator to teachers they should be counted and checked. When they

[11] Gerald S. Hanna, "Supervising Test Administration," Chapter 4 in Annie W. Ward, et. al (Editor), Guide for School Testing Program, National Council on Measurement in Education: (Washington, D.C., 1980), ERIC DOCUMENT 193 248

are received by the teacher, they should be counted. Similarly, as the test booklets are sent back, step by step, from the teacher to the district sorting facility, they should be counted. Why all this compulsive preoccupation with counting? The answer, of course, is twofold. First, counting test booklets, answer sheets, manuals for administering, etc., as they are successfully distributed from the director's office to the individual teacher, assures that clerical mistakes will not result in uncorrected shortages of materials. Being ready to administer a test to thirty students and discovering that one has only twenty-nine test booklets or answer sheets is not only awkward, but inexcusable. Second, counting tests at each stage of distribution and collection makes it relatively easy to pinpoint blame for missing tests. It has been this writer's experience that the ability to isolate blame for missing tests almost always prevents the loss of the tests in the first place.

6. Record Keeping

A basic requirement for a competency-based testing program is an effective means of maintaining and updating records of student performance. This is apparent when it is recognized that a teacher or administrator may have to determine quickly such information as (1) student competency weaknesses, (2) prescriptions for remediation, and (3) disposition of the corrective actions.

During the course of their school careers, students will be tested on many competencies, and profiles of performance will be unique for each. In order to keep track of competency performance and the needs for retesting, it is essential that a district have a carefully designed record-keeping system. In addition, a well designed competency-based testing program will require that systematic records be kept of the remediation which has been made, based on test performance.

The design of a record-keeping system should take into consideration the information and reporting needs of a particular program. Each district will need to define carefully the specific kinds of data to which reference will be made within a CBT program. It also will be necessary to identify the nature of the reporting which will be required. For example, is the system to have the capability of producing periodic paper reports of a predetermined nature? Likewise, should the system allow one to query the data base at any time to learn specific information about any given student? In short, the nature of the system will be a function of a variety of factors, including those suggested above.

While very simple paper and pencil records can be maintained, such procedures rarely will be adequate. The presence of microcomputer capabilities in nearly every district will allow

automated approaches to be developed. The complexity of the automated systems depend on several factors, among them are:

1. The number of students to be tracked.
2. The number of tests to be administered.
3. The number of competencies required.
4. Special policies for unique groups of students, such as students with exceptional needs or students who are not proficient in English.
5. Types and frequency of reporting requirements.
6. Complexity of the data requirements.
7. Frequency of updating.

While record keeping software of the type required for monitoring a CBT program is not widely available at this time, it can be expected that this kind of software will proliferate over the next several years. Thus, it is recommended that districts do a thorough search before embarking on software creation of their own. In cases where appropriate programs cannot be located, districts should consider cooperative development with other school districts.

Even though the needs for record keeping may seem remote during the beginning stages of CBT planning, it is important that this facet of the program receive early attention. The preparation for the design and creation of automated systems are nearly always complex and lengthy tasks.

7. Reporting

In principle, reporting the results of competency-based testing is no different from reporting the results of other objective-referenced testing programs. As suggested on pages 11-12, there are a variety of topics which should be addressed in any report. Above all, it needs to be remembered that different reports will have to be written for different audiences, which have varying needs and levels of understanding.

As in other testing programs, summary reports which show class, school, and district performance data will be of considerable value. However, the objective in competency testing is to assist each student so that he or she can demonstrate competency. It is for this reason that individual student reports are necessary.

If a computerized scoring service is used, it is advisable to purchase computer-generated individual student reports. The reports should be distributed to parents, with accompanying explanatory materials. At the least, there should be a basic description of the

competency program, a list of required competencies, and a definition of mastery.

8. Remediation

The goal of competency-based education is to raise the achievement levels of students who have failed to reach the competency levels. It is essential that students who do poorly be given as much help as is reasonable to allow them to master all competencies. (The goal, of course, is to bring each student to that level whereby he or she is able to master the established competencies).

Competency-based testing, of course, allows one to identify students who are failing to meet prescribed standards. Of even greater advantage is the fact that a competency-based testing program will allow for identification of specific learner strengths and weaknesses. Keep in mind, however, that identifying students who fail to meet prescribed standards represents only the initial step in any remedial program. Further assessment (diagnosis) of student strengths and weaknesses, formal observation by teachers, and conferences with students and their parents represent examples of additional steps which can be taken.

9. Evaluation

The final component of a competency-based testing program is a plan for evaluation.[12] There are a variety of purposes for CBT program evaluation; however, two would appear to be most significant:

1. Are the CBT program goals being attained?
2. What is the impact of CBT on the total school environment (e.g., are there unintended consequences)?

Evaluation should focus on the program goals, but also should consider the program as it is being implemented (referred to as summative evaluation).

Following are selected recommendations for developing an effective evaluation:

1. A comprehensive plan, with appropriate timelines, should be developed. In the process of developing that plan, representatives of important school and community groups should be involved.

[12] A valuable resource is *Handbook in Research and Evaluation* by Stephen Isaac and William B. Michael, Second Edition, Edits Publishers (San Diego: 1981).

2. The plan should be practical in its design, as determined by available resources.
3. Each of the nine elements identified in this section of the paper should be evaluated on a regular, on-going basis. If necessary, changes should be implemented.
4. A single individual (or small group of individuals) should be assigned primary responsibility for the evaluation process.
5. The evaluation plan should be especially sensitive to possible unexpected outcomes.

For more information about Wisconsin's competency-based testing legislation, contact the Department of Public Instruction.

PART III

TESTING IN WISCONSIN PUBLIC SCHOOL DISTRICTS

In January 1980, a questionnaire was sent to Wisconsin's 433 public school districts by the Wisconsin Pupil Assessment Program of the Department of Public Instruction. As noted in the introduction, the survey data is three years old. Nevertheless, the overall pattern found in 1980 is probably much the same. The questionnaire, completed by 303 (70%) districts, was designed to answer two questions about district testing programs:

1. What tests are used in Wisconsin public schools?
2. What use is made of test results?

Summary of Tests Used in Wisconsin Public School Districts

In response to the questionnaire, Wisconsin school districts indicated that tests currently used fall into three categories:[13]

1. Tests purchased from commerical testing companies
2. Tests purchased from the Wisconsin Pupil Assessment Program
3. District-developed tests

13It should be noted that there are a number of Wisconsin school districts use item collections offered by Cooperative Educational Service Agencies (CESA).

1. Commercial Tests

In terms of commerical tests, the questionnaire asked each district what tests were currently being used, or had been used during the previous three years.[14]

In addition to naming the tests used, districts were asked to indicate the grade levels of test use. The choices were K-5, 6-8, and 9-12. The results of this section of the questionnaire are shown in Table 1.

Besides showing the test name, test publisher, and grade usage, Table 1 indicates the type of test, either norm- or objective-referenced.[15] It is important not to equate the frequency of test use by districts, as shown in this table, with the number of students actually tested. For example, Milwaukee uses several commercial tests in its citywide testing program. One of these tests, the Differential Aptitude Tests, is shown in Table 1 to have a relatively modest frequency of use throughout the state. However, because it is used in the state's largest district, it may, in fact, be one of the tests taken by the greatest number of students in Wisconsin.

Two conclusions can be drawn from the data shown in Table 1. First, in nearly all cases, the commercial tests used by Wisconsin school districts are norm-referenced. Second, the commercial tests are produced by relatively few tests publishers, as shown in Table 2. In this case, the five most frequently used commercial tests for each of the three grade level categories is shown.

---

[14]Respondents also were asked to list tests used in their districts, but not appearing on the questionnaire. More than 150 additional commercial tests were listed. However, only six of these were reported ten or more times: Armed Services Vocational Aptitude Battery (16 times), Cognitive Abilities Test (14 times), Key Math Diagnostic Arithmetic Test (14 times), Peabody Individual Achievement Test (13 times), Wechsler Intelligence Scale for Children (10 times) and Woodcock Reading Mastery Tests (16 times).

[15]The distinction between norm-referenced and objective-referenced tests is discussed in Part I of this paper.

TABLE 1

COMMERCIAL TESTS USED IN WISCONSIN PUBLIC SCHOOL DISTRICTS

Name of Test	Publisher	Type of Test	Number of Districts Reporting Use K-5	6-8	9-12
Achievement Series: SRA Assessment Series	Science Research Associates	NRT	49	4	34
Analysis of Skills	Scholastic Testing Service	NRT	1	0	0
California Achievement Tests (CAT)	CTB/McGraw-Hill	NRT	32	26	18
Comprehensive Tests of Basic Skills (CTBS)	CTB/McGraw-Hill	NRT	18	23	24
Cooperative School and College Ability Tests	Addison-Wesley	NRT	1	2	17
Differential Aptitude Tests	Psychological Corp.	NRT	1	10	20
Gates-MacGinitie Reading Tests	Riverside Pub.	NRT	106	88	47
Henmon-Nelson Tests of Mental Ability	Riverside Pub.	NRT	43	57	95
Iowa Tests of Basic Skills (ITBS)	Riverside Pub.	NRT	89	88	24
Iowa Tests of Ed. Dev. (ITED)	Science Research Associates	NRT	3	5	32
Lorge-Thorndike Intel. Tests	Riverside Pub.	NRT	27	25	8
Metro. Achievement Tests	Psychological Corp.	NRT	79	39	20
Metropolitan Readiness Tests	Psychological Corp.	NRT	112	3	1
Objective-Referenced Bank of Item and Tests (ORBIT)	CTB/McGraw-Hill	ORT	1	1	0
Otis-Lennon Mental Ability Tests	Psychological Corp.	NRT	81	61	33
Prescriptive Read. Inven. (PRI)	CTB/McGraw-Hill	NRT	8	3	1
Sequential Tests of Educational Progress (STEP)	Addison-Wesley	NRT	15	33	78
Stanford Achievement Tests	Pshchological Corp.	NRT	84	73	23
Stanford Test of Academic Skills	Psychological Corp.	NRT	6	7	19
STS Educational Development Series	Scholastic Testing Service	NRT	6	7	10
Tests of Academic Progress	Riverside Pub.	NRT	5	2	15
School Curriculum Objective-Referenced Evaluation (SCORE)	Westinghouse Research Corp.	ORT	0	2	1
SRA Criterion-Referenced Measured Program	Science Research Associates	ORT	6	6	7
Stanford Diagnostic Math. Test	Psychological Corp.	NRT	21	19	8
Stanford Diagnostic Read. Test	Psychological Corp.	NRT	44	38	16
Wide Range Achievement Test	Guidance Assoc.	NRT	68	61	52

2. Tests Purchased from The Wisconsin Pupil Assessment Program

All tests developed by the Wisconsin Pupil Assessment Program are objective-referenced. Each has gone through a systematic developmental process. Committees consisting of teachers, parents,

TABLE 2

FIVE MOST FREQUENTLY USED COMMERCIAL TESTS

IN GRADE LEVELS K-5, 6-8, AND 9-12

Test	Publisher	Number of Districts Reporting Use
	Grades K-5	
Metropolitan Readiness Tests	Psychological Corporation	112
Gates-MacGinitie Reading Tests	Riverside Publishing	106
Iowa Tests of Basic Skills	Riverside Publishing	89
Stanford Achievement Tests	Psychological Corporation	84
Otis-Lennon Mental Ability Tests	Psychological Corporation	81
	Grades 6-8	
Iowa Tests of Basic Skills	Riverside Publishing	88
Gates-MacGinitie Reading Tests	Riverside Publishing	88
Stanford Achievement Tests	Psychological Corporation	73
Otis-Lennon Mental Ability Tests	Psychological Corporation	61
Wide Range Achievement Test	Guidance Associates	61
	Grades 9-12	
Henmon-Nelson Tests of Mental Ability	Riverside Publishing	95
Sequential Tests of Educational Progress	Addison-Wesley Publishers	78
Wide Range Achievement Test	Guidance Associates	52
Gates-MacGinitie Reading Tests	Riverside Publishing	47
Achievement Series: SRA Assessment Survey	Science Research Associates	34

employers, and others from throughout the state have assumed a central role in identifying the objectives or skills that students should know at various grade levels. In a similar manner, committees have played a major role in writing or selecting test questions to measure established objectives. To date, tests have been developed in mathematics, reading, geometry, writing, science, U. S. Government, and economics.

Although tests developed by the Wisconsin Pupil Assessment Program are designed for purposes of statewide assessment, each is made available to public and private schools through a local option testing service. The Department of Public Instruction provides materials at cost, including test booklets, answer sheets, test administration guides, and the opportunity to use a contracted scoring service.

Table 3 shows the pattern of local option use in public school districts from 1979 to 1983, based on district enrollment.

TABLE 3

NUMBER AND PERCENT OF PUBLIC SCHOOL DISTRICTS USING TESTS DEVELOPED BY THE WISCONSIN PUPIL ASSESSMENT PROGRAM, 1979-1983*

District Enrollment	No of Districts	No of Local Option Users 1979-83	% of Districts Using Local Option 1979-1983
10,000-plus	8	3	37.5
5,000-9,999	15	7	46.6
4,000-4,999	6	3	50.0
3,000-3,999	21	8	38.1
2,000-2,999	38	12	31.6
1,000-1,999	112	41	36.6
500- 999	113	50	44.3
under- 499	100	26	26.0

*For more information about the local option, contact the Wisconsin Pupil Assessment Program office, GEF 3, 125 South Webster Street, Madison, Wisconsin, 53702 (Telephone 608-266-3126).

3. District-Developed Tests

Most Wisconsin School Districts surveyed indicate dependence on commercially produced norm-referenced tests or Wisconsin Pupil Assessment Program objective-referenced tests. Relatively few districts (perhaps 10-15 percent) have developed their own test instruments for purposes of district-wide testing. The numbers and subject areas are shown in Table 4.

Generally, districts have developed tests in three areas: mathematics, reading, and writing. About 6 percent of districts responding listed tests in subjects other than these three. There are perhaps two basic reasons for the focus on mathematics, reading, and writing. First, reading, mathematics, and writing represent fundamental or basic skill areas. Thus, the emphasis given to these subjects is to be expected. Second, developing valid and reliable objective-referenced tests in mathematics, reading, and writing is generally an easier task than developing a test in such subjects as science or social studies.

District Use of Test Results

The survey results suggest that tests, regardless of source, tend to be used for essentially the same purpose. Generally, the most frequently stated use of test results is to diagnose student strengths and weaknesses.

TABLE 4

NUMBER OF PUBLIS SCHOOL DISTRICTS USING TESTS DEVELOPED BY THE DISTRICT

1978-1980

Subject Area	Grade Levels Tested K-5	6-8	9-12
Mathematics	39 (13%)*	51 (17%)	41 (14%)
Reading	30 (20%)	26 (9%)	17 (6%)
Writing	8 (3%)	12 (4%)	18 (6%)
Others	19 (6%)	19 (6%)	6 (2%)

*These figures refer to the percent of K-5, 6-8 and 9-12 districts responding to the survey.

Other frequent uses are to identify curricular strengths and weaknesses, to assist in guidance and counseling, and to place students. In contrast, evaluation of teachers is seldom reported as a use of test results.

Tests developed by the Wisconsin Pupil Assessment Program or private testing companies are also used in many instances to compare students to an outside group. District-developed tests, of course, cannot be used for this purpose. The specific breakdown of test usage is shown in Table 5.

TABLE 5

REPORTED USES OF TESTS IN WISCONSIN PUBLIC SCHOOL DISTRICTS

Use of Test Results*	Number of Districts Using Various Tests Commercial Tests	Wis. Pupil Assessment	District-Developed
Identification of curricular strengths and weaknesses	225	74	65
Promotion or graduation	15	2	22
Evaluation of teachers	7	1	1
Pupil placement	205	19	60
Diagnosis of students' strengths or weaknesses	253	54	74
General idea of how students compare with others	270	72	NA
Research	98	16	16
Guidance and Counseling	248	45	36

*Note that a given test may have multiple uses.

CONCLUSIONS AND RECOMMENDATIONS

School districts increasingly are being held accountable to demonsrate the effectiveness of the educational process. One direct consequence of the current demands for accountability is that Wisconsin public school districts are doing a great deal of testing. However, it is apparent from the survey results that there are numerous inconsistencies between the uses of tests and the purposes for which the tests were designed.

Responses to the survey indicate that the majority of Wisconsin public school districts purchase tests, rather than develop their own district-wide test instruments. Furthermore, the survey indicates that the purchased tests are overwhelmingly norm-referenced. Norm-referenced tests are most appropriate for purposes of comparing students to other students, and to some extent, for student placement, research, and guidance and counseling. The data show that many districts do use norm-referenced tests for these reasons.

At the same time, there are many instances of districts using norm-referenced tests for identification of curricular strengths and weaknesses, promotion or graduation, diagnosis of individual student strengths and weaknesses, and in a few cases for evaluation of teachers. These uses of norm-referenced tests are not recommended, essentially because norm-referenced tests were not designed for such purposes. Objective-referenced tests are more appropriate for such purposes.

The available data do not explain why Wisconsin public school districts have chosen to use commercially produced norm-referenced tests as they do. However, some possible explanations include the following: (1) the tests have been used for a number of years and, therefore, are familiar to decision-makers; (2) the tests are convenient to use because they require no time and effort for development, piloting, and revision; (3) the tests are available at a known cost; (4) the tests cover a wide range of content and grade levels; (5) the tests usually provide norms for a general comparison of local achievement; (6) the tests have a high reliability and small measurement error; and (7) there may be a misunderstanding of the essential differences between norm-referenced and objective-referenced tests.

In light of the survey results, it is apparent that school districts need to review what they are doing in the area of testing. This means a careful examination of the purposes for testing. It also means giving attention all components of the existing testing program, with a special look at whether or not tests presently being used provide the necessary information to make important educational decisions. If there are weaknesses in the testing program, these weaknesses should be corrected. In some instances, this may require a revamping of the existing program.

School districts have both the capacity and the responsibility to assume a leadership role in the areas of testing and accountability. This responsibility includes establishing a testing program that is carefully planned and implemented. It includes informing the community about the purposes for testing and design of the district's testing program. It also includes informing the community about the benefits and limitations of any testing program. A strong school system requires community support. The willingness of school districts to assume a leadership role in testing and accountability will enhance this support.

4000-3F37620-83